International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

FrauDetect: Deep Learning Based Credit Card Fraudulency Detection System

Shanmukha Priya Sreenidhi Appalabatla[1], Abothu Sharath Kumar[2], Adidamu Pramod Sai[3], Avula Sanjana[4], Dr. S. Satyanarayana[5]

[1,2,3,4] Student, Dept. of Computer Science and Engineering (AIML), Malla Reddy University, Hyderabad, India
[5] Professor, Dept. of Computer Science and Engineering (AIML), Malla Reddy University, Hyderabad, India

---

Abstract - Huge increase in the internet usage has been observed since last decade. It led to the emergence of services like e-commerce, tap and pay systems, online bill payment systems, etc. have proliferated and become more widely used. Due to various online payment options introduced by e-commerce and other numerous websites, the possibility of online fraud has risen drastically. Thus, due to an increase in fraud rates, research on analyzing and detecting fraud in online transactions has begun utilizing various machine learning techniques. The Deep Learning techniques viz., Convolutional Neural Network Architecture is used to detect credit card frauds in the proposed model. The Principal Component Analysis transformation gives a set of numerical input variables as output which are taken as the features to be considered. Due to confidentiality concerns, some of the original characteristics and background information about the data are not entirely disclosed. Features of credit card frauds must be chosen carefully as they play important role when deep learning techniques are used for credit card fraud detection. The TensorFlow and Keras are used for the development of the current proposed model. The proposed model aims to predict credit card frauds with 91.9% accuracy.

Key Words: Credit card, Fraudulent, Deep learning, Principal Component Analysis, TensorFlow, Keras

1. INTRODUCTION

The advancement in technology led to the emergence of services like e-commerce, tap and pay systems, online bill payment systems, etc. have proliferated and become more widely used. Due to various online payment options introduced by e-commerce and other numerous websites, the possibility of online fraud has risen drastically. This opened the doors to the research on analysing and detecting fraud in online transactions by applying various techniques.

The current project focusses on detecting credit card fraudulency using deep learning techniques and aims to predict credit card frauds with at-most accuracy.

To detect credit card frauds and to come up with a deep learning model which will accurately predict whether the credit card transaction is fraudulent or not is the aim the project. The scope of the proposed project is to develop an application that,

- Enables to classify credit card transactions into fraudulent and non-fraudulent.
- Based on the dataset containing the features and attributes related to credit card
- Determines accuracy in detecting frauds.

Because of the time constraint, limited parameters are considered for the study and only selected DL techniques are being used to cross check the accuracy given by our model.

2. RELATED WORKS

Dejan Varmedja [1] studied numerous machine learning algorithms and analysed them relating to credit card fraud detection systems. Multilayer perceptron is used (Artificial neural network) which consist of 4 hidden layers and relu activation functioned is used that is to avoid negative values and optimizer used is Adam for its best performance. It is observed that random forest yields the finest result in case of credit card fraud detection.

Changjun Jiang [2] suggested a method for fraud detection clustering the homogeneous historical transaction data ended up in aggregating transactions using sliding window strategy.

Sahil Dhankhad [3] has applied supervised machine learning algorithms on the real-world data set - a Novel

Approach Using Aggregation Strategy and Feedback Mechanism. Algorithms to implement a super classifier using ensemble learning are developed and are compared with the performance of supervised algorithms implementing super classifier. Out of ten machine learning algorithms implemented, Logistic Regression evolved as better option for predicting fraud transactions.

Rishikeshan O V, Sakala Sai Kiran et.al., [5] proposed an improved algorithm for credit card fraud detection. That is named as Naïve Bayes improved K-nearest Neighbour method (NBKNN). They have used a dataset on which they had applied the algorithms to identify the fraudulent transaction in the taken dataset.

Mohamad Zamini [6] purposed an unsupervised fraud detection method using autoencoder based clustering. The autoencoder is an auto associator neural network, used to lower the dimensionality, extract the useful features, and increase the efficiency of learning in a neural network. European dataset with 2,84,807 transactions is considered for experimentation which resulted in 0.024 as training loss, 0.027 as validation loss and the mean non-fraudulent data is 75% less than the mean of reconstructive error.

Shiyang Xuan [7] tried to check the best among two random forests. Random-tree-based random forest CART-based random forest. They use different random forest algorithms to train the behaviour features of normal and abnormal transactions and both of the algorithms are different in their base classifications and their performance on the Chinese e-commerce company. 91.96% accuracy is exhibited by random forest and 96.7% in CART-based random forest. Since the data used is from the B2C dataset many problems arrived such as unbalanced data. Hence, the algorithm can be improved.

2.1. Material and Methods

Method is a practical implementation of an approach. Detailed plan is provided that helps to keep researchers on track, making the process smooth, effective, and manageable. The proposed model follows CNN architecture with the help of TensorFlow and Keras.

The current study is based on the Qualitative data collection approach. Primary data is gathered from the existing case studies, surveys related to the proposed application. Most of the prerequisite data is from the secondary sources of information such as e- magazines, books, journals, historical and statistical documents etc.,

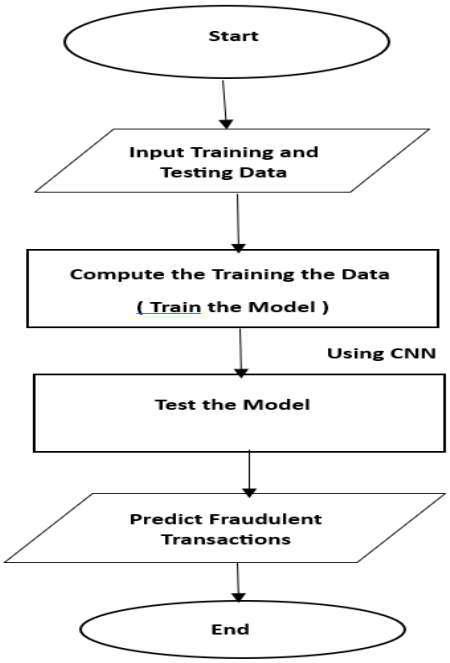

Fig 1: Flowchart representing the end-to-end process.

- The data is taken as input which is in the .csv file format.
- The dataset is divided into training and testing data (generally the training data size consists of 70-80% and rest is taken as testing data)
- The model is trained by the given training data. By undergoing multiple epochs, the model tries to decrease the loss function for exact prediction and detection.
- By the multiple layered architecture of CNN, it is easy to formulate and extract the features to compare with the data given and comes out with the total number of actual frauds, total number of

frauds detected, total number of non-frauds, total number of non-frauds predicted.

- At last, the accuracy of the model to detect the credit card frauds is obtained.

2.2. Dataset Description

For any model to understand how to perform various actions, training datasets must first be fed into the machine learning algorithm, followed by validation datasets (or testing datasets) which ensures that the interpretation of the model data is accurate.

Credit card fraud datasets are highly imbalanced, with a small fraction of transactions being fraudulent. Handling imbalanced data is a crucial challenge in fraud detection. Techniques like oversampling, under sampling, and cost-sensitive learning have been employed to address this issue.

The proposed project is carried over on the existing dataset. Credit card fraud detection datasets typically include a wide range of features that can help identify fraudulent transactions, while the specific features may vary depending on the dataset.

The dataset consists of multiple features like Transaction amount, Transaction type, Time of transaction, Merchant ID, Merchant country, Merchant state, Number of previous transactions, Average transaction amount, Number of transactions within a time window, Card usage patterns, Device information, Account age, Card expiration date and Card issuing bank so on. Each attribute has its own importance. The dataset used is relatively small as it contains only 30,000 transactions.

2.3. Data Pre-processing Techniques

First step in deep learning workflow is pre-processing data which ensures that the data is in a format that the network can accept. Before preparing the data to be fed to the network the input dataset must be pre-processed.

Sample data is obtained which further undergoes the cleaning step where the null, missing, and irrelevant data is handled. Then the data is split into respective testing, training, and validation sets. TensorFlow and Keras is used to create feedforward neural network. The model is evaluated using several metrics and most importantly the ROC AUC score and curve are obtained.

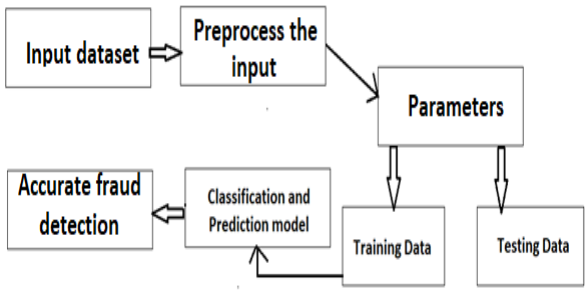

Fig 2: Data pre-processing flow

2.4. Building a Model

The model is built on several layers. They are as follows:

- Input Layer: It receives the input data, which is a 3D array of credit card transaction features.
- Convolutional Layers: They use filters to convolve over the input data to generate feature maps. The first convolutional layer has 32 filters with a kernel size of 3, and the second convolutional layer has 64 filters with a kernel size of 3. The output of the convolutional layers is passed to a max pooling layer.
- Max Pooling Layer: It down-samples the feature maps by taking the maximum value from each pool.
- Dense Layers: The output of the max pooling layer is passed to a fully connected dense layer with 128 units, which is then passed to another dense layer with a single output unit.
- Output Layer: It generates a single output value between 0 and 1, which represents the probability that the transaction is fraudulent.

2.5. Model Evaluation

The credit card fraud detection using Keras and TensorFlow with a convolutional neural network (CNN) architecture. The model on the testing set and prints the loss and accuracy scores. The model finally evaluates the following:

- Basic investigation on the input dataset.
- Number of fraudulent and non-fraudulent transactions.
- Total number of records.
- The accuracy of model to predict or detect credit card frauds after training.

2.6. Results and Discussion

Deep learning techniques, such as deep neural networks and recurrent neural networks, have gained attention in recent years for credit card fraud detection. Deep learning models can automatically learn hierarchical representations of data, enabling them to capture intricate fraud patterns. However, these models often require a large amount of labelled training data and substantial computational resources.

```
Epoch 1/100
2493/2493 [==============================] - 9s 3ms/step - loss: 0.0224
Epoch 2/100
2493/2493 [==============================] - 8s 3ms/step - loss: 0.0060
Epoch 3/100
2493/2493 [==============================] - 8s 3ms/step - loss: 0.0054
Epoch 4/100
2493/2493 [==============================] - 8s 3ms/step - loss: 0.0050
Epoch 5/100
2493/2493 [==============================] - 8s 3ms/step - loss: 0.0047
Epoch 6/100
2493/2493 [==============================] - 8s 3ms/step - loss: 0.0046
Epoch 7/100
2493/2493 [==============================] - 8s 3ms/step - loss: 0.0045
Epoch 8/100
2493/2493 [==============================] - 8s 3ms/step - loss: 0.0044
Epoch 9/100
2493/2493 [==============================] - 8s 3ms/step - loss: 0.0043
Epoch 10/100
```

Fig 3: Execution of Epochs

The V1, V2,……, V28 including Time and Amount are the features taken in the dataset. It is observed that there are no null or missing data in the dataset. The loss is calculated for 100 epochs on CNN model.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	39808
1	0.86	0.78	0.82	65
accuracy			1.00	39873
macro avg	0.93	0.89	0.91	39873
weighted avg	1.00	1.00	1.00	39873

Fig 4: Metrics before reshaping

The above values are obtained when the data is taken into consideration before reshaping.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	85300
1	0.81	0.84	0.82	143
accuracy			1.00	85443
macro avg	0.90	0.92	0.91	85443
weighted avg	1.00	1.00	1.00	85443

Fig 5: Metrics after reshaping

The above values are obtained when the data is taken into consideration after reshaping.

```
0    159207
1       284
Name: Class, dtype: int64
```

Class attribute contains 0, 1 label which represent fraudulent and non-fraudulent. It is observed that 1,59,207 non fraudulent credit card transactions took place while 284 fraudulent transactions occurred.

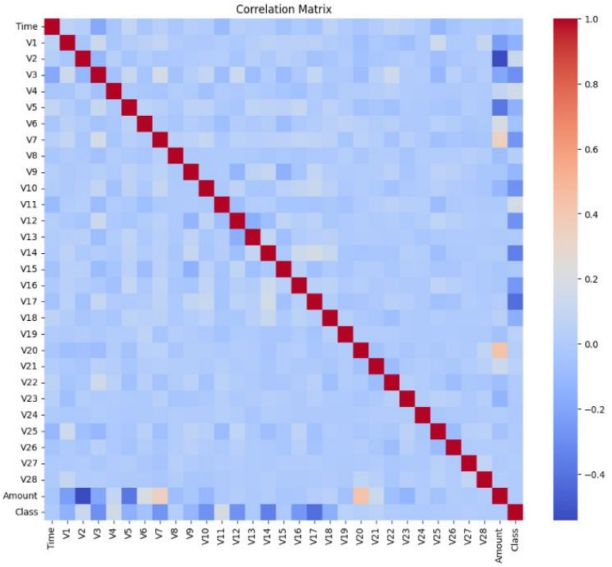

Fig 6: Correlation Matrix

. The correlation matrix of the dataset taken which is usually used to evaluate the correlation strength and patterns by finding out the linear relationships between variables and their potential dependencies

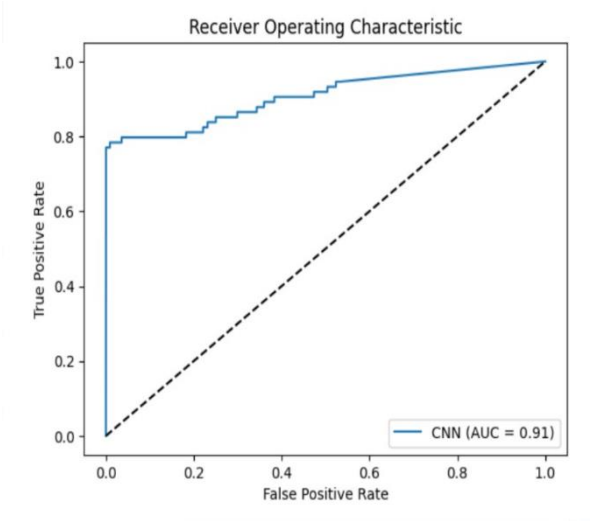

Fig 7: ROC AUC Curve

The ROC AUC Curve is obtained using which the overall performance of the model can be described. The AUC is 0.91 which is almost around the value 1 which depicts that the proposed model is an almost perfect classifier. The ROC AUC score value is around 91.9%.

```
roc_auc_score(test_targets, test_preds)
```

```
0.919410431303749
```

3. CONCLUSION

Credit card industry is most prone to fraudulency with the increase in popularity of electronic money transfers. To prevent these actions that lead to the drip of bank account information leak, skimming, counterfeit credit cards effectively and prevent the loss of reputation and customer loyalty, the implementation of advanced Credit Card Fraud Prevention and Fraud Detection methods should be considered by the credit card issuers.

A continuous improvement in the accuracy of fraud prevention based on each cardholder's behaviour information can be achieved by Deep Learning-based methods. It is very essential to train the Fraud Detection model continuously whenever new data is considered and arrives. Thus, new fraud patterns can be learned, and fraudulency can be detected as early as possible.

ACKNOWLEDGEMENT

We express our respect and profound gratitude to our Head of the Department, Dr. Thayyaba Khatoon Mohammed, who has been our inspiration in every step of our academic success. We thank our Mentor Dr. S. Satyanarayana for his encouragement and constant support.

REFERENCES

1. D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods," 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia, and Herzegovina, 2019, pp. 1-5, doi: 10.1109/INFOTEH.2019.8717766.
2. C. Jiang, J. Song, G. Liu, L. Zheng, and W. Luan, "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism," in IEEE Internet of Things Journal, vol. 5, no. 5, pp. 3637-3647, Oct. 2018, doi: 10.1109/JIOT.2018.2816007.
3. S. Dhankhad, E. Mohammed and B. Far, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study," 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 2018, pp. 122-125, doi: 10.1109/IRI.2018.00025.
4. K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," in IEEE Access, vol. 6, pp. 14277-14284, 2018, doi: 10.1109/ACCESS.2018.2806420.
5. Rishikeshan O V, Sakala Sai Kiran, Prasath S, Anitha M, "Credit Card Fraud Detection Using Isolation Forest and Local Outlier Factor", 2022, in International Journal of Scientific Research in Engineering and Management.
6. M. Zamini and G. Montazer, "Credit Card Fraud Detection using autoencoder based clustering," 2018 9th International Symposium on Telecommunications (IST), Tehran, Iran, 2018, pp.486-491,doi:10.1109/ISTEL.2018.8661129.

7. S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random Forest for credit card fraud detection," 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), Zhuhai, China, 2018, pp. 1-6, doi: 10.1109/ICNSC.2018.8361343.
8. Gupta, V, et.al., (2019). “Credit card fraud detection based on deep learning”, 2nd International Conference on Computer Science and Artificial Intelligence (CSAI) (pp. 255-260). IEEE. Dhaka.
9. Bhattacharyya, S., & Kalita, J. K. (2017) “Credit card fraud detection using Bayesian and neural networks”, Decision Support Systems, 75, 13-23.
10. Yin, X., Zhu, H., & Li, Q. (2016) “Credit card fraud detection using deep learning techniques”, 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1-5). IEEE.
11. Hu, X., Qi, H., Zhao, Y., & Hu, H. (2019) “Credit card fraud detection using deep learning with SMOTE and random undersampling”. In International Conference on Smart Grid and Electrical Automation (pp. 249-258). Springer.
12. Sudhakaran, S., & Shetty, N. (2018) “Deep learning for credit card fraud detection”. In Proceedings of the 2nd International Conference on Big Data, Cloud Computing, and Data Science (pp. 179-184).
13. Guo, S., & Zhang, D. (2018) “Deep learning for credit card fraud detection using autoencoders”. In 2018 15th Iranian Conference on Fuzzy Systems (IFSC) (pp. 1-5).
14. Ahn, G. J., Hu, H., & Kim, H. (2009) “Credit card fraud detection: A realistic modeling and a novel learning strategy”. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784-3797.